

From the INTERNATIONAL BUREAU

PCT

NOTICE INFORMING THE APPLICANT OF THE  
COMMUNICATION OF THE INTERNATIONAL  
APPLICATION TO THE DESIGNATED OFFICES

(PCT Rule 47.1(c), first sentence)

To:

ISHIDA, Takashi  
A. AOKI, ISHIDA & ASSOCIATES  
Toranomon 37 Mori Bldg., 5-1, Toranomon 3-chome  
Minato-ku, Tokyo 105-8423  
JAPON

98

Date of mailing(day/month/year)  
20 February 2003 (20.02.03)Applicant's or agent's file reference  
K869-PCT

## IMPORTANT NOTICE

International application No.  
PCT/JP02/007983International filing date(day/month/year)  
05 August 2002 (05.08.02)Priority date(day/month/year)  
03 August 2001 (03.08.01)

Applicant

FUJITSU LIMITED

1. Notice is hereby given that the International Bureau has **communicated**, as provided in Article 20, the international application to the following designated Offices on the date indicated above as the date of mailing of this notice:

US

In accordance with Rule 47.1(c), third sentence, those Offices will accept the present notice as conclusive evidence that the communication of the international application has duly taken place on the date of mailing indicated above and no copy of the international application is required to be furnished by the applicant to the designated Office(s).

2. The following designated Offices have waived the requirement for such a communication at this time:

EP, JP

The communication will be made to those Offices only upon their request. Furthermore, those Offices do not require the applicant to furnish a copy of the international application (Rule 49.1(a-bis)).

3. Enclosed with this notice is a copy of the international application as published by the International Bureau on 20 February 2003 (20.02.03) under No. WO 03/014966.

4. **TIME LIMITS for filing a demand for international preliminary examination and for entry into the national phase**

The applicable time limit for entering the national phase will, **subject to what is said in the following paragraph**, be **30 MONTHS** from the priority date, not only in respect of any elected Office if a demand for international preliminary examination is filed before the expiration of **19 months** from the priority date, but also in respect of any designated Office, in the absence of filing of such demand, where Article 22(1) as modified with effect from 1 April 2002 applies in respect of that designated Office. For further details, see *PCT Gazette* No. 44/2001 of 1 November 2001, pages 19926, 19932 and 19934, as well as the *PCT Newsletter*, October and November 2001 and February 2002 issues.

In practice, time limits other than the 30-month time limit will continue to apply, for various periods of time, in respect of certain designated or elected Offices. For **regular updates on the applicable time limits** (20, 21, 30 or 31 months, or other time limit), Office by Office, refer to the *PCT Gazette*, the *PCT Newsletter* and the *PCT Applicant's Guide*, Volume II, National Chapters, all available from WIPO's Internet site, at <http://www.wipo.int/pct/en/index.html>.

For filing a demand for international preliminary examination, see the *PCT Applicant's Guide*, Volume I/A, Chapter IX. Only an applicant who is a national or resident of a PCT Contracting State which is bound by Chapter II has the right to file a demand for international preliminary examination (at present, all PCT Contracting States are bound by Chapter II).

It is the applicant's sole responsibility to monitor all these time limits.

The International Bureau of WIPO  
34, chemin des Colombettes  
1211 Geneva 20, Switzerland

Authorized officer

Judith Zahra

Facsimile No.(41-22) 740.14.35

Telephone No.(41-22) 338.91.11

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
20 February 2003 (20.02.2003)

PCT

(10) International Publication Number  
WO 03/014966 A3

(51) International Patent Classification<sup>7</sup>: G06F 17/27,  
17/30

XU, Guowei [CN/CN]; C/O Fujitsu Research & Development Center Co., Ltd., Room B1003, Eagle Run Plaza No.26 Xiaoyun Road, Chaoyan District Beijing 100016, P.R. China (CN).

(21) International Application Number: PCT/JP02/07983

(22) International Filing Date: 5 August 2002 (05.08.2002)

(74) Agents: ISHIDA, Takashi et al.; A. AOKI, ISHIDA & ASSOCIATES, Toranomon 37 Mori Bldg., 5-1, Toranomon 3-chome, Minato-ku, Tokyo 105-8423 (JP).

(25) Filing Language: English

(26) Publication Language: English

(81) Designated States (*national*): JP, US.

(30) Priority Data:  
01123845.3 3 August 2001 (03.08.2001) CN

(84) Designated States (*regional*): European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR).

(71) Applicant (*for all designated States except US*): FUJITSU LIMITED [JP/JP]; 1-1, Kamikodanaka 4-chome, Naka-hara-ku, Kawasaki-shi, Kanagawa 211-8588 (JP).

Published:  
— with international search report.

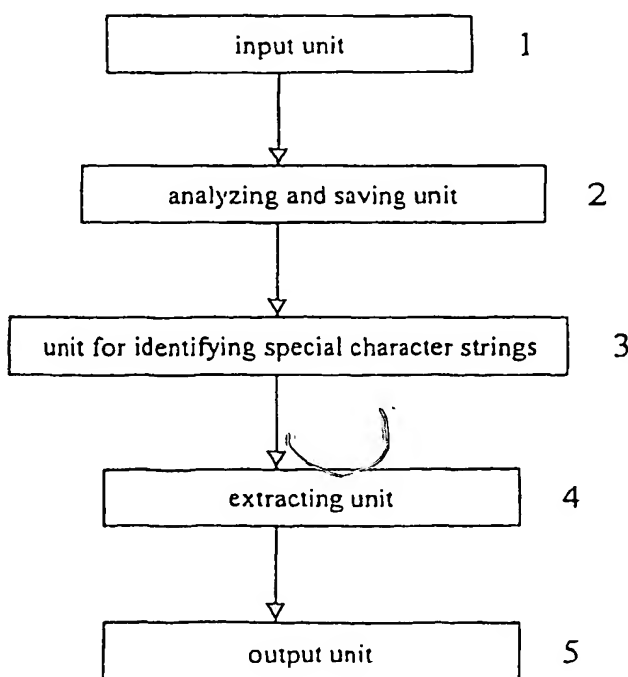
(72) Inventors; and

(88) Date of publication of the international search report:  
30 October 2003

(75) Inventors/Applicants (*for US only*): HUANG, Xiaohong [CN/CN]; C/O Fujitsu Research & Development Center Co., Ltd., Room B1003, Eagle Run Plaza No.26 Xiaoyun Road, Chaoyan District Beijing 100016, P.R. China (CN).

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: AN APPARATUS AND METHOD FOR EXTRACTING INFORMATION FROM A FORMATTED DOCUMENT



(57) Abstract: The present invention discloses an apparatus for extracting information from a formatted document, comprising: an input unit (1) for inputting a formatted document; a unit (2) for analyzing the input formatted document and saving the particular typographic information, a unit (3) for identifying special character strings on the basis of the analysis result by means of the typographic information such as font size, character font, color, etc.; a unit (4) for extracting the identified special character strings; and an output unit (5) for outputting the extracted character strings. When the typographic information of a certain character string is determined as a special typographic information, said character string is determined to be special character string. Thus, the present apparatus is able to automatically extract information from different types of format documents.

WO 03/014966 A3

## INTERNATIONAL SEARCH REPORT

Int. Patent Application No

PCT/JP 02/07983

A. CLASSIFICATION OF SUBJECT MATTER  
 IPC 7 G06F17/27 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, EPO-Internal, INSPEC, IBM-TDB, PAJ

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	WO 00 65483 A (SURFNOTES INC ;HIRSCH SCOTT (US)) 2 November 2000 (2000-11-02) page 15, line 18 -page 20, line 15	1-3,5, 7-9,11 4,6,10, 12
A	US 6 044 375 A (GREIG DARRYL ET AL) 28 March 2000 (2000-03-28) column 2, line 58 -column 3, line 48	1-12
A	ANONYMOUS: "Method of HTML Page maintenance" RESEARCH DISCLOSURE, no. 448, 1 August 2001 (2001-08-01), page 1394 XP002245253 Havant, UK, article No. 448120 page 1394	1,7
	--- -/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents:

\*A\* document defining the general state of the art which is not considered to be of particular relevance

\*E\* earlier document but published on or after the international filing date

\*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*8\* document member of the same patent family

Date of the actual completion of the international search

24 June 2003

Date of mailing of the international search report

07/07/2003

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Authorized officer

Bowler, A

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/JP 02/07983

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EMBLEY D W ET AL: "A conceptual-modeling approach to extracting data from the Web" BRIGHAM YOUNG UNIVERSITY, 1998, XP002181257 Provo, Utah Retrieved from the Internet: <URL:http://citeseer.nj.nec.com/24307.html > 'retrieved on 2001-10-25! page 8, paragraph 3.2 -page 17, paragraph 5 ---	1,7
A	"METHODOLOGY FOR SEARCHING ADOBE ACROBAT PORTABLE DATA FORMAT FILES BASED ON CONTENT RELEVANCE" RESEARCH DISCLOSURE, KENNETH MASON PUBLICATIONS, HAMPSHIRE, GB, no. 432, April 2000 (2000-04), page 756 XP000968936 ISSN: 0374-4353 page 756 ---	1,7
A	PATENT ABSTRACTS OF JAPAN vol. 2000, no. 02, 29 February 2000 (2000-02-29) & JP 11 328218 A (NIPPON TELEGR &AMP;TELEPH CORP &LT;NTT&GT;), 30 November 1999 (1999-11-30) abstract ---	1,7
P,X	US 6 298 357 B1 (YOUNG JEFFREY E ET AL) 2 October 2001 (2001-10-02) column 4, line 7 -column 4, line 21 -----	1,7

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/JP 02/07983

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 0065483	A	02-11-2000	AU 4493000 A WO 0065483 A2	10-11-2000 02-11-2000
US 6044375	A	28-03-2000	NONE	
JP 11328218	A	30-11-1999	NONE	
US 6298357	B1	02-10-2001	NONE	